Convergence of stochastic approximation algorithms with dependent random variables

AUTHOR(S):

渡辺, 正文

数理解析研究所講究録
611巻 1987年 187-200

Convergence of stochastic approximation algorihms with dependent random variables

福岡大 理 渡辺正文 (Masafumi Watanabe)

§1. 序

統計的構造がほとんど未知な系において最適解を推定する方法の一つとして確率近似法が1950年代より研究されて来ている。そのアルゴリズムは逐次的でかつ簡単であるため応用価値が高い。この報告は前回の研究集会の報告の続きであり特に，Robbins-Monro型の確率近似アルゴリズムのa.s.収束と平均収束に関するものである。ここでは，2r次平均収束を中心に議論する。従属確率変数列をもつ確率近似法の研究は，1970年代半ば頃より研究されている。(文献[1]〜[7]，[9]〜[11] etc.)。従属の場合における収束はほとんどがa.s.収束に関してであり，平均収束に関するものはあまりない，例えば，文献[1]，[2]，[8]〜[11]位である。独立の場合は両方の収束がほとんど同じ条件下で成立するのに対し，従属の場合は積のモーメントの評価がうまくいかぬため困難である。この報告に

おいては，平均収束に関しては従属性を3つのタイプのmixing条件で与える（strong-φ-mixing，φ-mixing，weak-φ-mixing これらの定義はStout [8] による）。またここで議論する確率近似法は応用上重要な線型の場合を含む条件の下で考察する。文献[1]においては弱い独立性（weak-φ-mixing）の下でa.s.収束と平均収束が論じられているが，線型の場合を含まぬ。一方，[2]では線型のみを扱っている。しかし，従属性はM-dependentを仮定している。

§2. Robbins-Monro 型確率近似法

R^N ； N-次元 Euclid 空間，内積 $\langle\cdot,\cdot\rangle$，ノルム $\|\cdot\|$

$\mathcal{B}^N$ ： R^N の Borel field

$(\Omega, \mathcal{A}, P)$ ；確率空間，以下考える確率変数は全てこの確率空間上で定義されているとする。

$\{\mathcal{A}_n\}$ ； $\mathcal{A}$ の sub-σ-fields の列，

$$\mathcal{A}_m^n \overset{\text{def.}}{=} \sigma\{\mathcal{A}_m, \mathcal{A}_{m+1}, \cdots, \mathcal{A}_n\}$$

$M(x)$ ： $R^N \longrightarrow R^N$，$\mathcal{B}^N$-可測

$Y_n(x,\omega)$ ； $R^N \times \Omega \longrightarrow R^N$，$\mathcal{B}^N \times \mathcal{A}_n$ −可測，$n=1,2,\cdots$，

$$Z_n(x,\omega) \equiv Y_n(x,\omega) - M(x), \quad n=1,2,\cdots$$

<u>Robbins-Monro stochastic approximation algorithm</u> ；

方程式 $M(x)=0$ の解（存在を仮定し，それを θ とする）

を観測列 $\{Y_n(\cdot,\omega)\}$ を用いて推定するアルゴリズムを次で与える，

$$\begin{cases} X_1(\omega) = R^n \text{の定数ベクトル（任意）} \\ X_{n+1}(\omega) = X_n(\omega) - a_n Y_n(X_n(\omega), \omega) \quad , \quad n=1,2,\cdots, \end{cases}$$

$$\left(= X_n(\omega) - a_n M(X_n(\omega)) - a_n Z_n(X_n(\omega), \omega) \right) ,$$

ここで，$\{a_n\}$ は正の単調減少列で

(A 1) $$\sum_{n=1}^{\infty} a_n^2 < \infty \quad , \quad \sum_{n=1}^{\infty} a_n = \infty$$

を満たすものとする。このとき，次の定理は一般的（$M(x)$ に関する仮定が最も一般的）なものとして知られている。

定理 2.1.（Gladyshev, E.G. : On Stochastic Approximation, Theory of Prob. and its appl. Vol.10, 1965, 275-278）．条件 (A.1) の他に，以下を仮定する。

(i) $$\inf_{\varepsilon < \|x-\theta\| < \varepsilon^{-1}} \langle x-\theta, M(x) \rangle > 0 \qquad \text{for each } \varepsilon > 0$$

(ii) $$\|M(x)\| \leq k(\|x\| + 1)$$

(iii) $$E[\, Y_n(X_n) \mid X_1, \cdots, X_n \,] = M(X_n) \quad \text{a.s.} \quad n=1,2,\cdots$$

(iv) $$E[\, \|Z_n(X_n)\|^2 \mid X_1, \cdots, X_n \,] \leq K(\|X_n\|^2 + 1), \ \text{a.s.} \quad n=1,2,\cdots$$

$$\Rightarrow \quad \lim_{n\to\infty} \|X_n - \theta\| = 0 \ \text{ a.s. かつ } \lim_{n\to\infty} E\|X_n - \theta\|^2 = 0$$

$\{\omega_n\}$ が独立の場合は本質的に (iii) が成立する。$\{\omega_n\}$ が独立でない場合，すなわち，(iii) が成立しない場合は上の条件の他にいくつかの条件が必要となる。この報告では $M(x)$ の条件は

出来る限り変えず，$M(x)$ 以外の条件を強めるまたは新たに加えることにより a.s. 収束と 2r 次平均収束について考察する。

§3. a.s. 収束

定理 3.1. 定理 2.1. の条件 (i), (ii) が成立。さらに，

(A2) $$\sup_n |a_n^{-1} - a_{n+1}^{-1}| < \infty$$

(B1) $^\exists\{\delta_n(\omega)\}$; 正の確率変数列

(i) $\delta_n(\omega) \downarrow 0$, $\sup_n \delta_n(\omega)\delta_{n+1}^{-1}(\omega) < \infty$

(ii) $\sup_n \delta_n(\omega) \|X_n(\omega)\| < \infty$

(B2) $^\exists\{\alpha_n(\omega)\}$; 非負確率変数列

(i) $\|Z_n(x,\omega)\| \leqq \alpha_n(\omega)(\|x\| + 1)$, $n = 1, 2, \cdots$

(ii) $\sum_{n=1}^{\infty} a_n^2 \alpha_n^2(\omega) < \infty$

(iii) $\sum_{n=1}^{\infty} a_n \delta_n(\omega) \alpha_n(\omega) < \infty$

(B3) $^\exists \gamma(\omega)$: 非負確率変数

(i) $\sup_n a_n \| \sum_{j=1}^{n} \delta_j^{-3} Z_j(x,\omega) \| \leqq \gamma(\omega)(\|x\| + 1)$

(ii) $\sup_n a_n \| \sum_{j=1}^{n} \delta_j^{-3}\{Z_j(x,\omega) - Z_j(y,\omega)\} \| \leqq \gamma(\omega) \|x - y\|$

$$\Rightarrow \quad \lim_{n\to\infty} \|X_n(\omega) - \theta\| = 0$$

（注）(1) 条件 (B1)(ii) は次の条件でおきかえることができる。

$^{\exists}\{\beta_n(\omega)\}$; 非負確率変数列

(i) $\langle \theta - x, Y_n(x,\omega) \rangle \leq \beta_n(\omega)(\|x\| + 1)$, $n = 1, 2, \cdots$

(ii) $\sum_{n=1}^{\infty} a_n \delta_n(\omega) \beta_n(\omega) < \infty$

この条件を満たす例として次の線型の例がある。

$M(x) = Ax + b$, A は $N \times N$-matrix , $b \in R^N$

$Y_n(x,\omega) = A_n(\omega)x + b_n(\omega)$, $A_n(\omega)$ は $N \times N$-random matrix, $b_n(\omega)$ は random vector , とすると, 条件 $\langle x, A_n(\omega)x \rangle \geq 0$ が成立する場合は, (B1)(ii)が満たされる。あるいは, 線型とは限らぬ場合, $\sup_{n,x} \|Z_n(x,\omega)\| \leq \delta(\omega)$ が成立する場合も(B1)(ii)が成立する([1]はこの場合を扱っている)。

(2) この定理においては確率的な議論は必要でない。確率論的議論は(B3)を導くときに必要となる。

§4. 2r次平均収束

$\mathcal{A}', \mathcal{A}''$ を $\mathcal{A}$ の sub-σ-fields とする。

$\varphi_1(\mathcal{A}', \mathcal{A}'') \overset{\text{def}}{=} \sup \{ \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right| ; P(A)P(B) \neq 0, A \in \mathcal{A}', B \in \mathcal{A}'' \}$

$\varphi_2(\mathcal{A}', \mathcal{A}'') \overset{\text{def}}{=} \sup \{ |P(B|A) - P(B)| ; P(A) \neq 0, A \in \mathcal{A}', B \in \mathcal{A}'' \}$

$\varphi_3(\mathcal{A}', \mathcal{A}'') \overset{\text{def}}{=} \sup \{ |P(A \cap B) - P(A)P(B)| ; A \in \mathcal{A}', B \in \mathcal{A}'' \}$

さらに、

$$\phi_1(n) \stackrel{def}{=} \sup_k \varphi_1(\mathcal{A}_1^k, \mathcal{A}_{k+n})$$

$$\phi_2(n) \stackrel{def}{=} \sup_k \varphi_2(\mathcal{A}_1^k, \mathcal{A}_{k+n})$$

$$\phi_3(n) \stackrel{def}{=} \sup_k \varphi_3(\mathcal{A}_1^k, \mathcal{A}_{k+n})$$

（注） (1) $\phi_1(n)$は任意のnに対して有限とは限らぬ。

(2) $\phi_1(n) \geqq \phi_2(n) \geqq \phi_3(n)$.

(3) $\lim_{n\to\infty} \phi_1(n) = 0$ のとき、$\{\mathcal{A}_n\}$は strong-ϕ-mixing

$\lim_{n\to\infty} \phi_2(n) = 0$ のとき、$\{\mathcal{A}_n\}$は ϕ-mixing

$\lim_{n\to\infty} \phi_3(n) = 0$ のとき、$\{\mathcal{A}_n\}$は weak-ϕ-mixing

という。

<u>補助定理 4.1.</u> $\mathcal{A}', \mathcal{A}''$を$\mathcal{A}$の sub-$\sigma$-fields. X, Yを各々$\mathcal{A}'$-可測，$\mathcal{A}''$-可測な確率変数とする。このとき，以下が成立.

(i) $E|X| < \infty$, $E|Y| < \infty$

$\Rightarrow$ $|EXY - EXEY| \leqq \varphi_1(\mathcal{A}', \mathcal{A}'')\, E|X|\, E|Y|$

(ii) $0 < p, q < 1$, $p^{-1} + q^{-1} = 1$, $E|X|^p < \infty$, $E|Y|^q < \infty$

$\Rightarrow$ $|EXY - EXEY| \leqq 2\varphi_2^{1/p}(\mathcal{A}', \mathcal{A}'')(E|X|^p)^{1/p}(E|Y|^q)^{1/q}$

(iii) $0 < p, q, s < 1$, $p^{-1} + q^{-1} + s^{-1} = 1$, $E|X|^q < \infty$,
$E|Y|^s < \infty$

$\Rightarrow$ $|EXY - EXEY| \leqq 10\varphi_3^{1/p}(\mathcal{A}', \mathcal{A}'')(E|X|^q)^{1/q}(E|Y|^s)^{1/s}$

(注) (3)は W. Phillip and W. Stout, "Almost sure invariance principles for partial sums of weakly dependent random variables", Mem. Amer. Math. Soc. 6 (1975). Lemma 7.2.1. による。 (i), (ii), (iii) は R^N の確率変数に対しても同様な結果が成立する。

補助定理 4.2. $\{X_n\}$: 確率変数列. t ∈ 正数とする.

(i) $\sup_n E|X_n|^t = c < \infty$

(ii) $X_n \longrightarrow X$ in probability

$\Rightarrow$ $\lim_{n\to\infty} E|X_n - X|^{t'} = 0$ for $t' < t$

(注) 補助定理 4.2. は Loéve, Probability Theory, p164 による。

補助定理 4.3. $\{\xi_n\}, \{u_n\}, \{v_n\}$: 非負実数列

$0 < t \leq 1$, $\sum_{n=1}^{\infty} v_n < \infty$,

$\xi_{n+1} \leq (1+v_n)\xi_n + u_n(\xi_n^{1-t} + 1)$, $n=1,2,\cdots$

$\Rightarrow$ $\xi_n^t \leq 2^t \left(\prod_{i=1}^{n} (1+v_i)^t \right) \left(\sum_{i=1}^{n} u_i \right)$, $n=1,2,\cdots$

(注) 証明は [11] lemma 1.

(A1), (A2)を§2及び§3で与えたものとする。さらに，以下の仮定を与える。以下，$0 < u \leq 1$，pは正整数とする。

(A3) $\exists \{\delta_n\}$: 正実数列

(i) $\delta_n \downarrow 0$, $\sup_n \delta_n \delta_{n+1}^{-1} < \infty$

(ii) $\sum_{n=1}^{\infty} a_n \delta_n^u < \infty$

(iii) $a_n \delta_n^{-(2p+1)} > a_{n+1} \delta_{n+1}^{-(2p+1)}$, $n=1,2,\cdots$

(C1) $\inf_{\varepsilon < \|x-\theta\|} < x-\theta, M(x) > \; > 0$ for each $\varepsilon > 0$

(C2) $\|M(x)\| \leq K(\|x\| + 1)$

(C3) $\exists F : R^N \to R^{N_0}$, Borel-可測，$F(\theta) = 0$，$\exists \{G_n(\omega)\}$; R^{N_0}-valued r.v.'s の列，各nに対し，G_nは$\mathcal{A}_n$-可測，

(i) $< x-\theta, Z_n(x,\omega) > \; = \langle F(x), G_n(\omega) \rangle_0$, $n=1,2,\cdots$

(ii) $\|F(x) - F(y)\|_0 \leq K(\|x\| + \|y\| + 1)\|x-y\|$

(C4) $\exists \{\alpha_n(\omega)\}$: 非負確率変数列，各nに対しα_nは$\mathcal{A}_n$-可測，

$\|Z_n(x,\omega)\| \leq \alpha_n(\omega)(\|x\| + 1)$, $n=1,2,\cdots$

(C5) $\exists \{\beta_n(\omega)\}$; 非負確率変数列，各nに対しβ_nは$\mathcal{A}_n$-可測，

$< \theta - x, Y_n(x,\omega) > \; \leq \beta_n(\omega)(\|x\|^{2-u} + 1)$, $n=1,2,\cdots$

(C6) $\sup_n a_n \delta_n^{-(2p+1)} \| \sum_{j=1}^{n} EG_j \|_0 < \infty$

(C7) $\exists M$（正整数）;

$$\sum_{n=M}^{\infty} a_n \delta_n^{-(2p+1)} \phi_1(n) < \infty$$

(C8) $\sup_n E[\gamma_n^{2p(1+M+\frac{2}{u})}] < \infty$,

ここで，$\gamma_n(\omega) = \max\{ \alpha_n(\omega), \beta_n(\omega), \|G_n(\omega)\|_0 \}$.

<u>補助定理 4.4.</u>　$\{V_n\}$：非負確率変数列，V_n は $\mathcal{A}_n$-可測。

$\{C_n\}$：非負実数列，以下を満たす。

(i)　$\sum_{n=1}^{\infty} C_n < \infty$,　$C_{n+1} \leq C_n \leq 1$

(ii)　$^{\exists}M$（正整数）：　$\phi_1(M) < \infty$

(iii)　$\sup_n E V_n^{\alpha M} < \infty$　for　$\alpha \geq 1$

$\Rightarrow$　$\sup_n E\left[\prod_{i=1}^{n} (1 + C_i V_i)^{\alpha}\right] < \infty$

(注)　証明は [11]，Lemma 2.

<u>定理 4.1.</u>　r を正整数とする。(A1)～(A3)，(C1)～(C8) が $p = r+1$ として成立。

$\Rightarrow$　$\lim_{n\to\infty} E\|X_n - \theta\|^{2r} = 0$

略証)　補助定理 4.3，4.4 を用いて，$\sup_n \delta_n^{\alpha} E\|X_n\|^{\alpha} < \infty$ が示される，ここで，$\alpha = 2p\left(1 + \frac{u}{Mu+2}\right)$。このことを用いて，補助定理 4.1 (i) より，$\sum_{n=1}^{\infty} a_n E[\|X_n\|^{2p-2} < F(X_n), G_n >.]$ 収束，を得る。この結果と，(C1) より，$\lim_{n\to\infty} E\|X_n - \theta\|^{2p}$ 存在，$^{\exists}\{X_{n_k}\} \subset \{X_n\}$，$\lim_{k\to\infty} \|X_{n_k} - \theta\| = 0$ a.s. を得る。従って，補助定理 4.2 より結論を得る。

次に以下の仮定を与える。

(C4)'　$^{\exists}\{\alpha_n(\omega)\}$；非負確率変数列，各 n に対して $\alpha_n(\omega)$ は

$\mathcal{A}_n$－可測，

$$\| Z_n(x,\omega) \| \leqq K (\|x\| + \alpha_n(\omega) + 1) \quad , \quad n=1,2,\cdots$$

(C7)′ $\quad \sum_{n=1}^{\infty} a_n \delta_n^{-(2p+1)} \{\phi_2(n)\}^{\frac{2p+1}{2p+2}} < \infty$

(C8)′ $\quad \sup_n E\, \gamma_n^{\frac{2p+2}{u}} < \infty$,

ここで，$\quad \gamma_n(\omega) = \max\{ \alpha_n(\omega), \beta_n(\omega), \|G_n(\omega)\|_0 \}$.

<u>定理4.2.</u> (A1)～(A3)，(C1)～(C3)，(C4)′，(C5)，(C6)，(C7)′，(C8)′が $p=r+1$ として成立．さらに，(C7)′は $p=r$ に対しても成立するとする。

$$\Longrightarrow \quad \lim_{n\to\infty} E\|X_n-\theta\|^{2r} = 0$$

略証）補助定理4.3を用いて，$\sup_n \delta_n^{2p} E\|X_n-\theta\|^{2p} < \infty$ が示される。このことを用いて，補助定理4.1(ii)より，

$\sum_{n=1}^{\infty} a_n E[\|X_n\|^{2p-2} \langle F(X_n), G_n\rangle_0]$ 収束，が成り立ち，(C1)を用いて，$\lim_{n\to\infty} E\| X_n-\theta\|^{2p}$ 存在 ($\sup_n E\|X_n-\theta\|^{2p} < \infty$)，$\exists \{X_{n_k}\} \subset \{X_n\}$，$\lim_{k\to\infty} \|X_{n_k}-\theta\| = 0$ a.s. を得る。従って，補助定理4.2より結論を得る。

次に，また仮定を与える。

(C7)″ $\quad \sum_{n=1}^{\infty} a_n \delta_n^{-(2p+1)} \{\phi_3(n)\}^{\frac{1}{2p+2}} < \infty$

<u>定理 4.3.</u> (A1) ~ (A3), (C1) ~ (C3), (C4)′, (C5), (C6), (C7)″, (C8)′ が $p = r+1$ に対して成立。

$$\Rightarrow \quad \lim_{n\to\infty} E\|X_n - \theta\|^{2r} = 0$$

略証) 定理 4.2. と同様にして、補助定理 4.1 (iii) を用いて示される。

§5. 平均収束の order

この章では、$a_n = an^{-1}$ $(a>0)$, $n=1,2,\cdots$ とする。さらに以下の仮定を与える。

(C1)′ $\exists \lambda > 0$; $\langle x-\theta, M(x)\rangle \geq \lambda \|x-\theta\|^2$

(C6)′ $\|E G_n\|_0 = 0$, $n=1,2,\cdots$

<u>定理 5.1.</u> (C1)′, (C2), (C3) ~ (C5), (C6)′, (C8) が $p = 2r$ として成立さらに、

$$\phi_1(n) \leq K n^{-b} \quad (b>0) \quad , n = M, M+1, \cdots$$

が成立する。

$$\Rightarrow \quad E\|X_n - \theta\|^{2r} \leqq \begin{cases} K n^{-t} (\log n)^2 & , 0 < t \leq 1 \\ K n^{-1} & , t > 1 \end{cases}$$

ここで、$t = \min\{2a\lambda, b\}$ 。

略証) 証明は、[10], [11] と同様の手法を用いて示される。

定理 5.2. (C1)′, (C2), (C3), (C4)′, (C5), (C6)′, (C8)′ が $p = r+1$ に対して成立，さらに，

$$\phi_2(n) \leqq K n^{-b} \quad (b>0), \quad n=1,2,\cdots$$

$\Rightarrow$ 定理 5.1 の結果が成立する，ただし，$t = \min\{2a\lambda, \frac{b(2r+1)}{2r+2}\}$。

定理 5.2. (C1)′, (C2), (C3), (C4)′, (C5). (C6)′, (C8)′ が $p = r+1$ に対して成立，さらに，

$$\phi_3(n) \leqq K n^{-b} \quad (b>0), \quad n=1,2,\cdots$$

$\Rightarrow$ 定理 5.1 の結果が成立する，ただし，$t = \min\{2a\lambda, \frac{b}{2r+2}\}$。

References

[1] Borodin, A. N.: A stochastic approximation procedure in the case of weakly dependent observations. Theory Prob. Appl. 24 (1979), 34–52.

[2] Eweda, E and Macchi, O.; Quadratic mean and almost-sure convergence of unbounded stochastic approximation algorithms with correlated observations. Ann. Institut. Henri Poincaré, Vol. 19, No. 1 (1983).

[3] Györfi, L. : Stochastic approximation from ergodic sample for linear regression. Z. Wahrsheinlichkeitstheorie verw. Geb. 54 (1980), 47-55.

[4] Kushner, H. J. and Clark, D. S. : Stochastic approximation methods for constrained and unconstrained systems. Springer 1978.

[5] Ljung, L. : Analysis of recursive stochastic algorithms. IEEE Trans. Automat. Contr. Vol. AC-22 (1977), 551 - 575.

[6] Ljung, L. : Strong convergence of a stochastic approximation algorithms. Ann. Statist. 6 (1978), 680 - 696.

[7] Métivier, M. and Priouret, P. : Application of a Kushner and Clark lemma to general classes of stochastic algorithms. IEEE Inform. Theory Vol. IT-30 (1984), 140 - 151.

[8] Stout, W. F. : Almost sure convergence. Academic Press, 1974.

[9] Watanabe, M. : A stochastic approximation from dependent observations. Z. Wahrsheinlichkeitstheorie verw. Geb. 62 (1983), 279 - 292.

[10] Watanabe, M. : The $2r$-th mean convergence of adaptive filters with stationary dependent random variables. IEEE Inform. Theory Vol. IT-30 (1984), 134–140.

[11] Watanabe, M. : The $2r$-th mean convergence of a stochastic approximation algorithm with weakly dependent observations. Fukuoka Univ. Science Reports vol. 16, No. 2 (1986), 71–80.